ST5210 Multivariate Data Analysis Cheatsheet

Yeo Ming Jie, Jonathan

Ch 3: Multivariate Distributions
Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. Then

$$\underline{Y} = A\underline{X} + \underline{b} \sim N_q\left(A\underline{\mu} + \underline{b}, A\Sigma A^T\right)$$

Any subset (q-component) of $\underline{X}$ is q-variate normal.
Theorem: Let $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p(\underline{\mu}, \Sigma)$ and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Let $\underline{X}_{2.1} = \underline{X_2} - \Sigma_{21}\Sigma_{11}^{-1}\underline{X}_1$. Then

$$\underline{X}_1 \sim N_r\left(\underline{\mu}_1, \Sigma_{11}\right) \text{ and } \underline{X}_{2.1} \sim N_{p-r}\left(\underline{\mu}_{2.1}, \Sigma_{22.1}\right)$$

where
$\underline{\mu}_{2.1} = \underline{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\underline{\mu}_1$ **and** $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Conditional Distribution of $\underline{x}_2$ given $\underline{x}_1 = \underline{x}$:

$$\left(\underline{X}_2 | \underline{X}_1 = \underline{x}_1\right) \sim N_{p-r}\left(\underline{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\underline{x}_1 - \underline{\mu}_1\right), \Sigma_{22.1}\right)$$

Theorem:
If $\underline{X}_1 \sim N_r\left(\underline{\mu}_1, \Sigma_{11}\right)$ and $\left(\underline{X}_2 \mid \underline{X}_1 = \underline{x}_1\right) \sim N_{p-r}\left(A\underline{x}_1 + \underline{b}, \Omega\right)$, where Ω does not depend on $\underline{x}_1$, then

$$\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix} \sim N_p(\underline{\mu}, \Sigma)$$

where

$$\underline{\mu} = \begin{pmatrix} \underline{\mu}_1 \\ A\underline{\mu_1} + \underline{b} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}A^T \\ A\Sigma_{11} & \Omega + A\Sigma_{11}A^T \end{pmatrix}$$

Central Limit Theorem:
If $\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_n$ are independent and identically distributed with mean $\underline{\mu}$ and variance matrix Σ, and $\underline{X} = \frac{1}{n}\sum_{i=1}^n \underline{X}_i$, then

$$\sqrt{n}(\underline{\bar{X}} - \underline{\mu}) \xrightarrow{\mathcal{L}} N_p(\underline{0}, \Sigma) \quad \text{for } n \to \infty$$

Note: No assumption of normality required.
Transformation of Statistics:
If $\sqrt{n}(\underline{W} - \underline{\mu}) \xrightarrow{\mathcal{L}} N_p(\underline{0}, \Sigma)$ and if

$$\underline{g}(\underline{W}) = \left(g_1(\underline{W}), \cdots, g_q(\underline{W})\right)^T : \mathbb{R}^p \to \mathbb{R}^q$$

are real valued functions differentiable at $\underline{\mu} \in \mathbb{R}^p$, then

$$\sqrt{n}(\underline{g}(\underline{W}) - \underline{g}(\underline{\mu})) \xrightarrow{\mathcal{L}} N_q\left(\underline{0}, D^T\Sigma D\right) \quad \text{for } n \to \infty$$

where $D = \left(\left(\frac{\partial g_j}{\partial w_i}\right)(\underline{W})\Big|_{\underline{W}=\underline{\mu}}\right)$ is a $(p \times q)$ matrix of all partial derivatives.

Ch 4: Sampling Distributions
Let $\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_n$ be a random sample taken from a multivariate normal distribution $N_p(\underline{\mu}, \Sigma)$.

1. (a) $\underline{\bar{X}} \sim N_p\left(\underline{\mu}, \frac{1}{n}\Sigma\right)$
   (b) $(n-1)S \sim W(n-1, \Sigma)$
   (c) $\underline{\bar{X}}$ and S are independent
2. $n(\underline{\bar{X}} - \underline{\mu})^T\Sigma^{-1}(\underline{\bar{X}} - \underline{\mu}) \sim \chi^2(p)$
3. $n(\underline{\bar{X}} - \underline{\mu})^T S^{-1}(\underline{\bar{X}} - \underline{\mu}) \sim T^2(p, n-1)$

$$\text{where } T^2(p, n-1) = \frac{(n-1)p}{n-p}F(p, n-p)$$

Ch 5: Hypothesis Testing
Want to test $\mathrm{H}_0 : \underline{\mu} = \underline{\mu}_0$ against $\mathrm{H}_1 : \underline{\mu} \neq \underline{\mu}_0$
Test Problem 1 (Σ known):

$$-2\log\lambda = n\left(\underline{\bar{X}} - \underline{\mu}_0\right)^T \Sigma^{-1}\left(\underline{\bar{X}} - \underline{\mu}_0\right) \sim \chi^2(p)$$

Reject for

$$-2\log\lambda > \chi^2_{0.95}(p)$$

Test Problem 2 (Σ unknown): Rejection region:

$$\left(\underline{\bar{X}} - \underline{\mu}_0\right)^T S^{-1}\left(\underline{\bar{X}} - \underline{\mu}_0\right) > \frac{(n-1)p}{n(n-p)}F_{1-\alpha}(p, n-p)$$

Confidence region for $\underline{\mu}$:

$$\left\{\underline{\mu} : (\underline{\bar{x}} - \underline{\mu})^T S^{-1}(\underline{\bar{x}} - \underline{\mu}) \leq \frac{(n-1)p}{n(n-p)}F_{1-\alpha}(p, n-p)\right\}$$

Simultaneous Confidence Intervals for $\underline{a}^T\underline{\mu}$:

$$\left|\frac{\sqrt{n}\left(\underline{a}^T\underline{\bar{X}} - \underline{a}^T\underline{\mu}\right)}{\sqrt{\underline{a}^T S\underline{a}}}\right| \leq t_{1-\alpha/2}(n-1)$$

or equivalently

$$\underline{a}^T\underline{\bar{X}} - \sqrt{K_\alpha \underline{a}^T S\underline{a}} \leq \underline{a}^T\underline{\mu} \leq \underline{a}^T\underline{\bar{X}} + \sqrt{K_\alpha \underline{a}^T S\underline{a}}$$

where $K_\alpha = \frac{(n-1)p}{n(n-p)}F_{1-\alpha}(p, n-p)$.

Ch 6: Two Group Analysis
Paired Samples (Sample size n small):
Test Statistic:

$$T^2 = n\underline{\bar{D}}^T \boldsymbol{S}_d^{-1}\underline{\bar{D}}$$

Reject H_0 if

$$T^2 > \frac{p(n-1)}{n-p}F_\alpha(p, n-p)$$

$100(1-\alpha)$% Confidence Region for $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ is

$$\left\{\underline{\delta} : n(\underline{\bar{D}} - \underline{\delta})^T S_d^{-1}(\underline{\bar{D}} - \underline{\delta}) \leq \frac{p(n-1)}{n-p}F_\alpha(p, n-p)\right\}$$

$100(1-\alpha)$% Bonferroni Simult. CIs for δ_j 's are:

$$\bar{D}_j \pm t_{\alpha^*/2}(n-1)\sqrt{\frac{S_{jj}}{n}}$$

where $\alpha^* = \alpha/p$ and S_{jj} is j-th diagonal entry of $\boldsymbol{S}_d$.
Paired Samples (Large sample size):

$$T^2 = n\underline{\bar{D}}^T \boldsymbol{S}_d^{-1}\underline{\bar{D}} \sim \chi^2(p) \quad \text{approximately}$$

Reject H_0 if $T^2 > \chi^2_\alpha(p)$.
$100(1-\alpha)$% Confidence Region for $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$:

$$CR = \left\{\underline{\delta} : n(D - \underline{\delta})^T S_d^{-1}(\underline{\bar{D}} - \underline{\delta}) \leq \chi^2_\alpha(p)\right\}$$

$100(1-\alpha)$% Bonferroni Simult. CIs for δ_j 's are:

$$\bar{D}_j \pm z_{\alpha^*/2}\sqrt{\frac{S_{jj}}{n}}$$

where $\alpha^* = \alpha/p$ and S_{jj} is the j^{th} diag entry of $\boldsymbol{S}_d$.
2 Indep. Samples (When n_1 or n_2 small):
Assume $\Sigma_1 = \Sigma_2$. Define

$$S_{pool} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1 + n_2 - 2}$$

Then

$$T^2 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)^T\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S_{\text{pool}}\right]^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

Reject H_0 if

$$T^2 > \frac{(n_1 + n_2 - 2)\,p}{n_1 + n_2 - 1 - p}F_\alpha\left(p, n_1 + n_2 - 1 - p\right) = \mathbf{k}$$

$100(1-\alpha)$% Confidence Region for $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ is

$$\left\{\underline{\delta} : (\underline{\bar{X}_1} - \underline{\bar{X}_2} - \underline{\delta})^T\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pool}}\right]^{-1}(\underline{\bar{X}_1} - \underline{\bar{X}_2} - \underline{\delta}) \leq \mathbf{k}\right\}$$

$100(1-\alpha)$% Bonferroni Simult. CIs for δ_j's are:

$$\left(\underline{\bar{X}_1} - \underline{\bar{X}_2}\right)_j \pm t_{\alpha^*/2}\left(n_1 + n_2 - 2\right)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S_{jj}}$$

where S_{jj} = j-th diagonal of S_{pool} and $\alpha^* = \alpha/p$.
2 Indep. Samples (Large sample size):

$$T^2 = \left(\underline{X}_1 - \underline{X}_2\right)^T\left[\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)\right]^{-1}\left(\underline{X}_1 - \underline{X}_2\right) \sim \chi^2(p)$$

Reject H_0 if the observed $T^2 > \chi^2_\alpha(p)$.
$100(1-\alpha)$% Bonferroni Simult. CIs for δ_j's:

$$\left(\bar{X}_1 - \underline{X}_2\right)_j \pm z_{\alpha^*/2}\sqrt{\frac{S^1_{jj}}{n_1} + \frac{S^2_{jj}}{n_2}}$$

where S^g_{jj} = j-th diagonal of $\boldsymbol{S}_g$ and $\alpha^* = \alpha/p$

MANOVA
Have $\mathbf{SSCP}_{\text{tot}} = \mathbf{SSCP}_{\text{trt}} + \mathbf{SSCP}_{\text{res}}$ where

$$\boldsymbol{SSCP}_{\text{tot}} = \sum_{g=1}^a n_g\left(\underline{\bar{X}}_{g.} - \underline{\bar{X}}_{..}\right)\left(\underline{\bar{X}}_{g.} - \underline{\bar{X}}_{..}\right)^T + \sum_{g=1}^a\sum_{i=1}^{n_g}\left(\underline{X}_{gi} - \underline{\bar{X}}_g\right)\left(\underline{X}_{gi} - \underline{\bar{X}}_{g.}\right)^T$$

and

$$\boldsymbol{SSCP}_{\text{res}} = (n_1 - 1)\boldsymbol{S}_1 + \cdots + (n_a - 1)\boldsymbol{S}_a$$

where

$$\boldsymbol{S}_g = \sum_{i=1}^{n_g}\left(\underline{X}_{gi} - \underline{\bar{X}}_g\right)\left(\underline{X}_{gi} - \underline{\bar{X}}_g\right)^T, \quad g = 1, \cdots, a$$

Test statistic

$$\Lambda = \frac{|\boldsymbol{SSCP}_{\text{res}}|}{|\boldsymbol{SSCP}_{\text{trt}} + \boldsymbol{SSCP}_{\text{res}}|}$$

Ch 7: Principal Component Analysis
Dimensionality Reduction. Define the j-th PC by

$$Y_j = \underline{u}_j^T\underline{X}$$

where $\underline{u}_j$ = eigenvector corresponding to λ_j. Note that $Var(Y_j) = \lambda_j$ and $Cov(Y_j, Y_t) = 0 \ (j \neq t)$, i.e. PCs are uncorrelated.
Note: Proportion of Total Variance Explained

$$PVE = \frac{Var(Y_j)}{tr(S)} = \frac{\lambda_j}{\sum_t \lambda_t}$$

PCA results dependent on scaling of variables; higher loading placed on variables with high variability. Preferred to work with standardized variables:

$$Z_i = \frac{X_i - \bar{X}}{\sqrt{\sigma_{ii}}} \quad \text{and} \quad R = V^{-1/2}\Sigma V^{-1/2}$$

where $V^{-1/2} = diag(\sigma_{11}^{-1/2}, \ldots, \sigma_{pp}^{-1/2})$.

Ch 8: Factor Analysis
Assume $E(\underline{F}) = \underline{0}$, $\mathrm{Var}(\underline{F}) = I_m$. (i.e. Factors are uncorrelated with mean 0 and variance 1. Let $E(\underline{U}) = \underline{0}$ where $\mathrm{Var}(\underline{U}) = \Psi = \mathrm{diag}\left(\psi_1, \cdots, \psi_p\right)$ and $\mathrm{Cov}(\underline{F}, \underline{U}) = 0_{m\times p}$.
Factor Analysis Model:

$$\begin{aligned} \underline{X} - \underline{\mu} &= \Lambda\underline{F} + \underline{U} \\ \underline{\mu} &= \text{ mean vector} \\ \Lambda &= (p \times m) \text{ loadings} \\ \underline{F} &= (m \times 1) \text{ common factors} \\ \underline{U} &= (p \times 1) \text{ specific factors} \end{aligned}$$

where $\underline{F} = (F_1, \cdots, F_m)^T, \underline{U} = (\epsilon_1, \cdots, \epsilon_p)^T$, and (j,k)-th entry of Λ is $\lambda_{j,k}$.
Taking $Var(X) = \Sigma$, then

$$\Sigma = \Lambda\Lambda^T + \Psi$$

Theorem:

$$\sigma_{jj} = \sum_{k=1}^m \lambda_{jk}^2 + \psi_j = h_j^2 + \psi_j$$

where $h_j^2 = \sum_{k=1}^m \lambda_{jk}^2$ is called the j-th communality, ψ_j is called the specific variance or uniqueness.
PVE: The k-th factor explains $d_k / \sum_{j=1}^p \sigma_{jj}$ of the total variance, where

$$d_k = \sum_{j=1}^p \lambda_{jk}^2$$

i.e. square sum of column entries (assuming column is the factor loadings)
Note: Preferred to work with standardized data, equivalent to ***using the sample correlation matrix*** R instead of sample covariance matrix S. Then, for each variable X_i,

$$(\text{Factor } 1)^2 + (\text{Factor } 2)^2 = \text{Communality}, h_j^2$$

and

$$\text{Specific Var}, \psi_j = 1 - \text{Communality}$$

Large Sample Test for No. of Common Factors:
It can be shown that

$$\chi^2 = \left(n - 1 - \frac{2p + 4m + 5}{6}\right)\log\left(\frac{\left|\widehat{\Lambda}\widehat{\Lambda}^T + \widehat{\Psi}\right|}{|S_n|}\right) \sim \chi^2(q),$$

where $q = \frac{1}{2}\left((p-m)^2 - p - m\right)$, where m denotes no. of factors, p dimension of the data.

Ch 9: Classification

Minimum ECM Classification Rule: Assign x_0 to π_t if

$$\sum_{k=1}^{m} p_k p(t|k) c(t|k)$$

is the smallest among all such sums (i.e. for different classes $t = 1, \cdots, m$)

When m = 2: Assign x_0 to π_1 if

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}$$

assign to π_2 otherwise.

Two Multivariate Normal Populations:

(a) Equal prior and equal misclassification costs:

Case 1: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}$

Under this assumption, we assign x_0 to π_1 only if

$$d_1\left(\underline{x}_0\right) < d_2\left(\underline{x}_0\right)$$

where and

$$d_i(\underline{x}) = \left(\underline{x} - \bar{x}_i\right)^T \hat{\Sigma}^{-1}\left(\underline{x} - \underline{\bar{x}}_i\right)$$

$$\hat{\Sigma} = S_{\text{pool}} = \frac{\left(n_1 - 1\right) S_1 + \left(n_2 - 1\right) S_2}{n_1 + n_2 - 2}$$

Equivalently, we may rewrite $d_1\left(\underline{x}_0\right) < d_2\left(\underline{x}_0\right)$ as $d_{12}\left(\underline{x}_0\right) > 0$ where $d_{12}\left(\underline{x}_0\right)$ given by

$$\left(\underline{\underline{x}}_1 - \underline{\bar{x}}_2\right)^T S_{\text{pool}}^{-1}\, \underline{x}_0 - \frac{1}{2}\left(\underline{\bar{x}}_1 - \underline{\bar{x}}_2\right)^T S_{\text{pool}}^{-1}\left(\underline{\bar{x}}_1 + \underline{\bar{x}}_2\right)$$

Case 2: $\boldsymbol{\Sigma_1} \neq \boldsymbol{\Sigma_2}$ (Use S_1 and S_2)

Assign $\underline{x}_0$ to π_1 if

$$d_1\left(\underline{x}_0\right) + \log\left|S_1\right| < d_2\left(\underline{x}_0\right) + \log\left|S_2\right|$$

where $d_i\left(\underline{x}_0\right) = \left(\underline{x}_0 - \underline{x}_i\right)^T S_i^{-1}\left(\underline{x}_0 - \underline{x}_i\right)$ for $i = 1, 2$.

(b) General Prior and Misclassification Costs:

Case 1: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}$

Assign $\underline{x}_0$ to π_1 if

$$d_{12}\left(\underline{x}_0\right) > k = \log\left(\frac{p2\ c(1|2)}{p1\ c(2|1)}\right)$$

where $d_{12}(x_0)$ given by

$$\left(\underline{\mu}_1 - \underline{\mu}_2\right)^T \Sigma^{-1} \underline{x}_0 - \frac{1}{2}\left(\underline{\mu}_1 - \underline{\mu}_2\right)^T \Sigma^{-1}\left(\underline{\mu}_1 + \underline{\mu}_2\right)$$

where

$$\hat{p}_1 = \frac{n_1}{n_1 + n_2} \quad \text{and} \quad \hat{p}_2 = \frac{n_2}{n_1 + n_2}$$

Case 2: $\boldsymbol{\Sigma_1} \neq \boldsymbol{\Sigma_2}$

Assign $\underline{x}_0$ to π_1 if

$$\frac{1}{2}\left\{\left[d_2\left(\underline{x}_0\right) - d_1\left(\underline{x}_0\right)\right] + \log\frac{\left|S_2\right|}{\left|S_1\right|}\right\} > k$$

where $d_i(x) = \left(\underline{x} - \underline{x}_i\right)^T S_i^{-1}\left(\underline{x} - \underline{x}_i\right)$ for $i = 1, 2$, and $k = \log\left(\frac{p_2 c(1|2)}{p_1 c(2|1)}\right)$